rchandra@utexas.edu
Google Scholar

Rohan Chandra
Homepage

Research Statement

Deploying intelligent mobile robots in the real world has been a longstanding goal in robotics and AI. Despite significant progress in perception, planning, and control, robots today fail to blend seamlessly, safely, and confidently into human environments (Figure 1).

(a) Driving in chaotic traffic (b) Navigating airports (c) Home robots

Figure 1: My research goal is to develop human-like mobility enabling robots to safely and efficiently navigate these environments.

Each scenario requires robots to jointly make sense of the chaos in the scene, reason about human behavior, and navigate safely, efficiently, and smoothly. For instance, navigating through airports requires agility to smoothly maneuver around dense pedestrian clusters, autonomous driving requires handling unknown environments and unpredictable on-road agents, and home robots have to deal with constrained spaces, pets, and children. Humans, on the other hand, can navigate these scenarios with ease by jointly **understanding the complex scene**, **reasoning about the behaviors of other agents**, and **planning safe, deadlock-free, and agile trajectories, even taking calculated risks when necessary**. **My research studies the differences between robot mobility and human-like mobility, and develops algorithms and systems to bridge this gap.**

⊙ **Previous and ongoing work**–Towards understanding complex scenes, I have proposed new methods for *(i)* **detection and tracking** of pedestrians and vehicles in dense and heterogeneous traffic [1, 2], *(ii)* **predicting their future trajectories** [3, 4, 5], *(iii)* **semantic segmentation** in adverse conditions [6, 7]. Ongoing work includes *(iv)* **action recognition** from RGB videos. Next, I have also developed online, robust, and general **agent behavior models** [8, 9, 10] for characterizing the aggressiveness or conservativeness of mobility in dense and heterogeneous environments. Finally, I introduced a novel bi-level optimization navigation framework algorithm for **risk-aware navigation** [11], **dynamic autonomous intersection control** [12, 13], **multi-agent pathfinding** [14], and **social navigation** [15, 16]. Lastly, I also developed simulators [17, 18], tools [4], and datasets [19] towards developing human-like mobility.

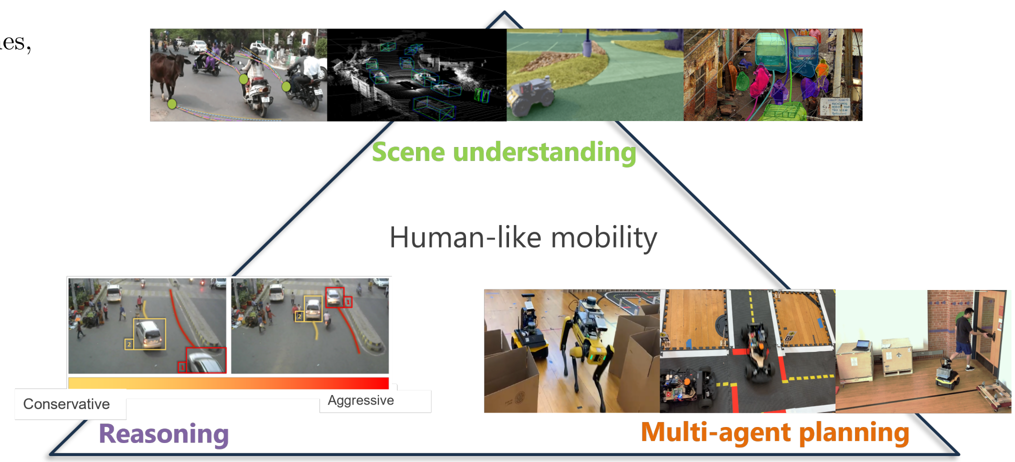

Figure 2: The three principles that define *Human-like Mobility*.

⊙ **Research Agenda**–During the initial years, my lab will conduct research along the following three directions: *(i) Everything Perception* for challenging environments–we will develop foundational models for a new framework towards integrated **multi-modal, multi-task**, and **risk-aware reasoning and intent inference**, *(ii)* **Deployable planning and control frameworks** for fully decentralized multi-agent navigation with generalizability, liveness, agility, and safety guarantees, and *(iii)* **autonomous racing** in unstructured environments.

⊙ **Impact**–Featured in science news and media such as TechXplore, my work has also been presented at several reputed workshops such as the UMD Future Faculty Fellows 2021, RSS Pioneers 2022, Microsoft Future Leader in Robotics and AI 2023, and KAUST Rising Star in AI 2023. My work also received the 2023 SNU Ph.D. Talk award, 2022 Charles A. Caramello Distinguished Dissertation finalist award and the UMD 2020 summer research fellowship. Based on my work, I have also organized three workshops on autonomous driving, multi-robot planning and interaction, and social navigation at IROS 2022, RSS 2023, and IROS 2023.

1 Thesis and Postdoc Work: Perception, Behavior Modeling, and Planning in Multi-agent Environments

1.1 Anticipating Human Motion in Crowded and Chaotic Environments

Human-like mobility enables robots to proactively navigate crowded and chaotic environments around other humans *by anticipating* their motion. Towards building a model of human understanding, robots must be able to accurately, robustly, and in realtime, track humans in these environments, estimate their risk tolerance, and predict their future motion. Not only are these tasks inherently challenging, but also the difficulty rises in dense and chaotic environments.

Tracking in Crowded Settings–To estimate risk tolerances of humans and predict their future motion, robots need to track the humans. In crowded and dynamic environments, however, humans are frequently occluded and tracks fragment. In **DensePeds** [1](IROS'19), I design and give a proof of correctness for a tracking-by-detection algorithm that segments out background noise from an agent's bounding box, leading to cleaner and more reliable tracks. Not only did my approach rank 1st on the MOT16 benchmark leaderboard [20] at the time, but it formally explains the success of Multi-Object Tracking and Segmentation (MOTS [21]), a now well known framework popularly used in the computer vision community. **RoadTrack** [2](ICRA'20) extends DensePeds by incorporating a motion model that simultaneously takes into account collision avoidance and inter-agent interactions to bolster tracks that are at high risk of fragmenting. My tracking algorithm has been adopted in navigation systems to deploy robots in crowded indoor building on campus [22, 23].

Motion forecasting in dense and dynamic environments–Forecasting human motion enables robots to effectively plan and navigate in human environments, but motion forecasting is challenging in crowded heterogeneous scenarios. The high density causes agents to execute highly non-linear trajectories that are non-trivial to predict. Furthermore, the predictions implicitly depend on the dynamics of an agent, which vary significantly in heterogeneous areas. In **TraPHic** [3](CVPR'19), I develop an algorithm that selectively focuses attention on fewer agents by training the ego-vehicle to identify the agents that deserve more importance than others. For instance, a pedestrian in the way of the ego-vehicle requires more attention than, say, a parked car to the side. TraPHic additionally takes in the explicit shape, size, and dynamics, of vehicles thereby allowing it to generate predictions for the ego-vehicle that reflect the traffic around it.

Classical deep trajectory forecasting is a supervised learning problem and hence depends on large training datasets that require time and effort to clean and process. In a follow up work **RobustTP** [4](ACM CSCS'19), I propose an end-to-end approach that does not require manually labeled ground-truth trajectories to train the trajectory prediction network. The input to this algorithm consists only of raw traffic videos obtained from commodity sensors such as monocular RGB cameras.

1.2 Online Human Risk Estimation in Crowded Environments

Dynamic environments necessitate risk estimation models to be run on data streaming in real time, capable of adapting to rapid pace of change in the environments, and above all, explainable when probed. The current paradigm of learning human objectives from data, widely accepted in the robotics community, fails in all three aspects. My thesis introduces an online, simple, and explainable framework for estimating humans' risk preferences that exploits the high density of traffic and crowds. It is robust enough to operate on *raw data* streamed directly from a camera or lidar and generalizes to traffic conditions in *four countries*, India, Singapore, China, and USA, with unique environments and drivers. Lastly, it requires no assumptions or dependency other than the last few seconds of a driver's trajectory. The key insight on which this framework rests is that a motorcyclist does not need to know the exact risk preference of the trucker next to him, the motorcyclist just has to know whether the trucker's appetite for risk is more or less than his own.

The algorithm, called **CMetric** [8, 10](IROS'20, IEEE ITS'21), observes raw vehicle trajectories and uses tools from graph theory such as vertex centrality functions to measure the likelihood and intensity of driving styles such as overspeeding, overtaking, sudden lane-changes, etc. I evaluate CMetric by measuring the time difference between the moments when a human identifies an aggressive behavior and when CMetric identifies the same behavior. Our experiments showed that CMetric can identify different behaviors with an average time difference of less than 0.02 seconds. A key step in our graph-theoretic algorithm is to compute the leading eigenvector of the Laplacian matrix of the traffic graphs, which can be an expensive operation as the number of traffic participants grow in crowded traffic. In follow up work [9], I recursively exploit sub-matrix information as well as exploit the sparsity and symmetry of Laplacian matrices to compute inverse Laplacian matrices more efficiently, thereby decreasing the time (by a factor of $2\times$) and space complexity of the eigenvalue algorithm.

The CMetric algorithm can be used for trajectory planning [11]. The planner takes into account the wide range of human driver behaviors on the road, from aggressive maneuvers like speeding and overtaking, to conservative traits like driving slowly and conforming to the right-most lane. Specifically, the planner learns a mapping from CMetric's output to a driver's entropic risk preference. The planner shows that in a merging scenario, the final trajectories obtained from the risk-aware planner generate desirable emergent behaviors.

A natural question arose: can we collectively predict low-level information such as trajectories and high-level information such as driver behavior from a single neural network? In **SpectralLSTM** [5](RAL/IROS'20), we propose a unique two-stream neural network where the first stream employs the TraPHic algorithm while the second stream in parallel channels the CMetric algorithm to predict the driver's behavior. This work also introduces a new regularization technique to stabilize training for trajectory prediction architectures. Lastly, I also derive a theoretical upper bound on the prediction error of the regularized trajectory predictions.

1.3 Multi-Agent Coordination and Planning in Constrained Spaces

In crowded, dynamic, and chaotic environments, robots' conservative nature causes them to frequently freeze in constrained spaces with multiple conflicting agents, which I term as *social mini-games*, in the interest of safety, resulting in jerky motion that is far from time-efficient. During my Ph.D. and postdoc, I have designed algorithms and systems for robots that can be safe, but also be less conservative, so that they can proactively resolve the freezing robot problem in social mini-games in a distributed or decentralized manner relying only on local sensor information.

Multi-agent coordination in social mini-games–One of the common causes for the freezing robot problem is when multiple agents end up in conflict in constrained, but crowded, environments *e.g.* multiple vehicles arriving at a four-way, unsignaled intersection at approximately the same time. Resolving such conflicts is hard because of inherent symmetry in the environment and that every robot wants to "win" the conflict. In **GamePlan** [12](RAL/ICRA'22), I introduce a new class of auctions that efficiently determine a fair ordering based on the priorities of each agent, which is determined using the CMetric algorithm. I prove that these auctions not only optimize each agent's specific utility but also maximize the global welfare of the system.

Applying coordination strategies to multi-agent planning–In SocialMAPF [14], I show that the auction-based coordination algorithm can be applied to discrete path planning problems. Specifically, I extend the multi-agent path finding (MAPF) problem for non-cooperative agents. Approaches for MAPF with non-cooperative agents are either optimal or efficient, but not both. I propose the first optimal *and* efficient solver for strategic agents. In GameOpt [13], I apply auction-based coordination in the continuous space domain towards cooperative intersection control for dynamic, multi-lane, unsignalized intersections. My approach combines an auction mechanism to generate a priority entrance sequence for every agent, with an optimization-based trajectory planner satisfying the priority sequence. This coupling operates at real-time speeds of less than 10 milliseconds in high density traffic of more than $10,000$ vehicles/hr, $100\times$ faster than other fully optimization-based methods, while providing guarantees in terms of fairness, safety, and efficiency. Tested on the SUMO simulator, my algorithm improves throughput by at least 25%, time taken to reach the goal by 75%, and fuel consumption by 33% compared to auction-based approaches and signaled approaches using traffic-lights and stop signs.

The auction framework indicated that game theory was the right tool to resolve deadlocks in social mini-games, but it still required a central auction master. In ongoing work [15, 16], I introduce a new class of decentralized controllers that ensure both safety and liveness by attaining a game-theoretic Nash equilibrium. I show that these controllers can be generally tacked on to any constrained optimization-based local trajectory planner such as model predictive control or dynamic window approach, by simply adding our controller as an additional constraint. My approach to ensuring liveness rests on two novel insights: (*i*) there exists a mixed-strategy Nash equilibrium that allows decentralized robots to perturb their state onto *liveness sets* i.e. states where robots are deadlock-free and (*ii*) forward invariance of liveness sets can be achieved identical to how control barrier functions (CBFs) guarantee forward invariance of safety sets. We successfully deploy[1] the proposed algorithm in the real human environments using F1/10 robots, a Clearpath Jackal, and a Boston Dynamics Spot as well as in simulated social mini-games. We show that (*i*) classical navigation performs far better than learning-based algorithms for multi-agent social robot navigation in terms of success rate, (*ii*) a controller obeying game-theoretic safety certificates is necessary for decentralized multi-robot social navigation, (*iii*) our approach is more socially compliant, that is, results in the fewest changes in velocities by up to $5\times$ and yields a flow rate of $2.0 - 2.1$ $(\text{ms})^{-1}$ which is comparable to flow rate in human navigation at 2.0 $(\text{ms})^{-1}$.

1.4 Tools and Resources

During my PhD, I released a software framework, TrackNPred [4] that benchmarks state-of-the-art tracking and trajectory prediction methods on real-world dense traffic datasets. I also introduced the METEOR dataset [19] a new traffic dataset of Indian traffic videos focusing on rare and interesting multi-agent driving behaviors, categorized into traffic violations, atypical interactions, and diverse scenarios. Finally, I present a new reinforcement learning simulation environment [17] by enriching existing traffic simulators with behavior-rich trajectories corresponding to varying levels of aggressiveness using CMetric.

In my postdoc, I developed SocialGym 2.0 [18], a simulator for multi-agent navigation research. Our simulator enables navigation for multiple autonomous agents, replicating real-world dynamics in complex indoor environments, including doorways, hallways, intersections, and roundabouts. Unlike current simulators that concentrate on single robots in open spaces, SocialGym 2.0 employs multi-agent reinforcement learning (MARL) to develop optimal navigation policies for multiple robots with diverse, dynamic constraints in complex environments.

[1] Watch video summary at https://youtu.be/fA7BbM8iTwg.

2 Research Agenda: Analyzing complex environments, Language-guided reasoning, and Integrated perception and control

2.1 Understanding Complex Environments

Foundational perception models for complex environments–While current perception algorithms demonstrate impressive performance in controlled or sparse settings, they often falter when faced with the complexities of real-world environments, which include diverse terrains, fluctuating weather conditions, and an intricate mix of dynamic and static obstacles. Given this, much active research has focused on enhancing individual perception tasks for these challenging contexts. My research direction, however, aims to break new ground by developing foundational models for an innovative framework called *Everything Perception*. This framework seeks to integrate multiple perception tasks, thereby facilitating robust multi-task perception in complex and unpredictable environments. As illustrated in my prior work [1], a strategically combined, albeit manually configured, approach of object detection and instance segmentation yielded superior tracking performance in dense traffic conditions. By further advancing these foundational models for Everything Perception, we aim to create a unified, adaptive algorithmic approach capable of navigating the multi-layered complexities of real-world settings.

LLM-driven introspective agents–In unfamiliar or dynamically changing environments, safety is contingent on a robot's ability to continuously monitor various factors such as weather conditions, time of day, terrain type, and nearby obstacles. Traditional methods often require manual intervention to correct failures, presenting a bottleneck in scalability and real-time adaptability. One promising approach to overcome this limitation is the concept of "introspective perception" [24] which enables robots to autonomously account for uncertainties in their environment. My research aims to leverage Large Language Models (LLMs) in developing advanced agents capable of introspective perception. Specifically, I am interested in investigating whether LLMs can be employed to create agents that autonomously quantify environmental uncertainties. The ultimate goal is to develop a systematic, multi-step algorithm where the introspective agent first evaluates static environmental elements, such as the terrain, followed by dynamic variables, like moving obstacles. Each step would involve a quantitative assessment of uncertainty, which can be aggregated to form a comprehensive safety index for robotic navigation.

2.2 Inferring and communicating intent

Representing human intent as a language–In India, drivers honk, not only to signal emergencies, but also to communicate intent, request coordination, assert dominance, convey frustration, among many other reasons. In short, honking is used as a means of communication during driving. Similarly, pedestrians often communicate intent via gaze or body language. In light of the progress in natural language understanding driven by large language models (LLMs), I am excited to investigate leveraging language as a means to directly represent these social cues. Although work has been done to study social cues such as gaze, body language, vehicle honking, and so on, my intuition is that these cues can be cast in the form of a language. I will use the expressive power of vision language models, built on top of LLMs, to build a language representation of social cues in order to learn humans' risk tolerances simply from realtime camera input without the need for lower level computer vision. I will also use LLMs as a means to study and analyze the impact of different cultural norms on behavior.

2.3 Integrated learning, perception, and control for human-like mobility

Human-like mobility constitutes many moving parts including perception, coordination, planning, and control. So far, I have established principles of human-like mobility in each of these separately. But for deployment in chaotic multi-agent environments, it is necessary to have a single seamless end-to-end systems.

A *G.L.A.S half full* approach to multi-agent navigation—*G.L.A.S half full* represents multi-agent navigation with **G**eneralizability, **L**iveness, **A**gility, and **S**afety guarantees in a **full**y decentralized setting. This approach mirrors the intricate choreography of human movement, capturing its essence in algorithmic form. At the heart of this paradigm lies a sophisticated blend of reinforcement learning and transfer learning techniques, empowering agents to adapt effortlessly to a kaleidoscope of environments. To maintain the decentralized nature and liveness of the system, I integrate cutting-edge communication, coordination, and deadlock resolution strategies. I will employ barrier functions to create an impenetrable shield against collisions and other hazards. Agility is achieved through optimal control algorithms, enabling high-speed navigation that dynamically responds to ever-changing conditions. What sets this paradigm apart is its interdisciplinary approach. Achieving these objectives necessitates an intellectual deep dive through adjacent domains, including machine learning, game theory, optimization, and control theory. The resulting algorithms are not mere patchworks but elegant systems that weave elements from each domain into a unified whole, fulfilling all objectives.

Towards stable deep neural multi-agent control–Multi-agent reinforcement learning (MARL) is a promising approach for decentralized planning and control in multi-agent settings. It also serves as a viable candidate as a building block for multi-agent planning and control in a potential end-to-end model for human-like mobility. However, MARL-based navigation policies have been used in a very limited sense due to the inherent instability in training them. Recently, in [25], we discovered that incorporating human intent into the training scheme stabilized training of decentralized MARL approaches achieving SOTA performance in autonomous driving in heterogeneous traffic. I am most exited to continue this line of research and leverage human intent and risk preferences to build new stable MARL algorithms for decentralized multi-agent navigation.

References

[1] R. Chandra, U. Bhattacharya, A. Bera, and D. Manocha, "Densepeds: Pedestrian tracking in dense crowds using front-rvo and sparse features," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 468–475, IEEE, 2019.

[2] R. Chandra, U. Bhattacharya, T. Randhavane, A. Bera, and D. Manocha, "Roadtrack: Realtime tracking of road agents in dense and heterogeneous environments," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1270–1277, IEEE, 2020.

[3] **Chandra, Rohan**, U. Bhattacharya, A. Bera, and D. Manocha, "Traphic: Trajectory prediction in dense and heterogeneous traffic using weighted interactions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8483–8492, 2019.

[4] **Chandra, Rohan**, U. Bhattacharya, C. Roncal, A. Bera, and D. Manocha, "Robusttp: End-to-end trajectory prediction for heterogeneous road-agents in dense traffic with noisy sensor inputs," in *ACM Computer Science in Cars Symposium*, pp. 1–9, 2019.

[5] **Chandra, Rohan**, T. Guan, S. Panuganti, T. Mittal, U. Bhattacharya, A. Bera, and D. Manocha, "Forecasting trajectory and behavior of road-agents using spectral clustering in graph-lstms," *IEEE Robotics and Automation Letters*, 2020.

[6] D. Kothandaraman, **Chandra, Rohan**, and D. Manocha, "Ss-sfda: Self-supervised source-free domain adaptation for road segmentation in hazardous environments," *arXiv preprint arXiv:2012.08939*, 2020.

[7] D. Kothandaraman, **Chandra, Rohan**, and D. Manocha, "Bomuda: Boundless multi-source domain adaptive segmentation in unconstrained environments," *arXiv preprint arXiv:2010.03523*, 2020.

[8] R. Chandra, U. Bhattacharya, T. Mittal, A. Bera, and D. Manocha, "Cmetric: A driving behavior measure using centrality functions," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2035–2042, IEEE, 2020.

[9] R. Chandra, U. Bhattacharya, T. Mittal, X. Li, A. Bera, and D. Manocha, "Graphrqi: Classifying driver behaviors using graph spectrums," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4350–4357, IEEE, 2020.

[10] **Chandra, Rohan**, A. Bera, and D. Manocha, "Using graph-theoretic machine learning to predict human driver behavior," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 2572–2585, 2021.

[11] **Chandra, Rohan**, M. Wang, M. Schwager, and D. Manocha, "Game-theoretic planning for autonomous driving among risk-aware human drivers," in *2022 International Conference on Robotics and Automation (ICRA)*, pp. 2876–2883, IEEE, 2022.

[12] **Chandra, Rohan** and D. Manocha, "Gameplan: Game-theoretic multi-agent planning with human drivers at intersections, roundabouts, and merging," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2676–2683, 2022.

[13] N. Suriyarachchi, **Chandra, Rohan**, J. S. Baras, and D. Manocha, "Gameopt: Optimal real-time multi-agent planning and control for dynamic intersections," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, pp. 2599–2606, IEEE, 2022.

[14] **Chandra, Rohan**, R. Maligi, A. Anantula, and J. Biswas, "Socialmapf: Optimal and efficient multi-agent path finding with strategic agents for social navigation," *IEEE Robotics and Automation Letters*, 2023.

[15] R. Chandra, R. Menon, Z. Sprague, A. Anantula, and J. Biswas, "Decentralized social navigation with non-cooperative robots via bi-level optimization," *arXiv preprint arXiv:2306.08815*, 2023.

[16] R. Chandra, V. Zinage, E. Bakolas, J. Biswas, and P. Stone, "Decentralized multi-robot social navigation in constrained environments via game-theoretic control barrier functions," *arXiv preprint arXiv:2308.10966*, 2023.

[17] A. Mavrogiannis, R. Chandra, and D. Manocha, "B-gap: Behavior-rich simulation and navigation for autonomous driving," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4718–4725, 2022.

[18] Z. Sprague, **Chandra, Rohan**, J. Holtz, and J. Biswas, "Socialgym 2.0: Simulator for multi-agent social robot navigation in shared human spaces," *arXiv preprint arXiv:2303.05584*, 2023.

[19] **Chandra, Rohan**, X. Wang, M. Mahajan, R. Kala, R. Palugulla, C. Naidu, A. Jain, and D. Manocha, "Meteor: A dense, heterogeneous, and unstructured traffic dataset with rare behaviors," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 9169–9175, IEEE, 2023.

[20] A. Milan, L. Leal-Taixé, I. Reid, S. Roth, and K. Schindler, "Mot16: A benchmark for multi-object tracking," *arXiv preprint arXiv:1603.00831*, 2016.

[21] P. Voigtlaender, M. Krause, A. Osep, J. Luiten, B. B. G. Sekar, A. Geiger, and B. Leibe, "Mots: Multi-object tracking and segmentation," in *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, pp. 7942–7951, 2019.

[22] A. J. Sathyamoorthy, J. Liang, U. Patel, T. Guan, **Chandra, Rohan**, and D. Manocha, "Densecavoid: Real-time navigation in dense crowds using anticipatory behaviors," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11345–11352, IEEE, 2020.

[23] A. J. Sathyamoorthy, U. Patel, T. Guan, and D. Manocha, "Frozone: Freezing-free, pedestrian-friendly navigation in human crowds," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 4352–4359, 2020.

[24] S. Rabiee and J. Biswas, "Introspective perception for mobile robots," *arXiv preprint arXiv:2306.16698*, 2023.

[25] X. Wu, **Chandra, Rohan**, T. Guan, A. S. Bedi, and D. Manocha, "iplan: Intent-aware planning in heterogeneous traffic via distributed multi-agent reinforcement learning," *arXiv preprint arXiv:2306.06236*, 2023.